M7.9 THREE DIMENSIONAL SUB-BAND CODING OF VIDEO

Gunnar Karlsson and Martin Vetterli

Department of Electrical Engineering
and
Center for Telecommunications Research
Columbia University, New York, NY 10027

ABSTRACT

A novel coding scheme has been developed which is based on multi-dimensional sub-band coding. In our method the digital video signal is filtered and sub-sampled in all three dimensions (temporally, horizontally and vertically) to yield the sub-bands, from which the input signal can be losslessly reconstructed in the absence of coding loss. The sub-bands can be more efficiently coded than the input signal in terms of compression and quality, because the restricted information in each band allows well-tailored encoding. The computational complexity of this coding scheme compares favorably to DCT with interframe prediction and inter/intra frame DPCM. The scheme has an architectural structure suitable for parallel implementation, and it yields high compression with sustained good quality.

1. INTRODUCTION

Sub-band coding was initially developed as a technique for speech compression. The underlying theory was extended to multi-dimensional signals [5] and the technique has since then been successfully used to compress images as well [6, 2]. We now propose three dimensional sub-band coding as a viable technique for video compression. It is believed to be a simple video coding method, with performance comparable to other methods, such as transform coding, intra/interframe DPCM, and vector quantization. In our method the video signal is analyzed into sub-sampled temporal and spatial frequency bands which are more tractable for coding than the input signal. This allows the compression to be adjusted according to perceptual criteria by making it highest in the frequency bands where the distortion become least visible. Note that coding schemes which operate on sub-blocks of the image may yield "blocking effects" (i.e., a visible block structure) when the blocks of pixels or transform coefficients are coarsely quantized. Moreover, redundancy due to correlation across block boundaries is not removed. In sub-band coding, the images are processed in their entirety wherefore the aforementioned problems do not exist. Furthermore, the sub-band coding scheme offers a low computational complexity and an algorithmic structure which lends itself to parallel implementation.

The presented work is aimed at video transmission over packet-switched networks. Since the bit rate of a compressed video signal varies according to the activity of the captured scene, packet-switching, which allows variable transmission rates, seems a suitable technique for video transmission. Thereby it is expected that the received video signal would have a constant perceptual quality. However, there are issues, normally not addressed in a pure coding context, which do have to be considered while designing the coder for packet-switched networks. An outline of pertinent issues, such as robustness to transmission with packet loss, appears in [3], where it was found that three dimensional sub-band coding offers a good framework to tackle these issues.

This paper expands on the coding method outlined in [3]. In section 2, we brief the theory of sub-band analysis and synthesis and give the implementation of the three dimensional analysis system. The encoding of the sub-bands is discussed in section 3. The computational complexity and the simplicity of parallel implementation are compared to discrete cosine transform coding and DPCM in section 4. The scheme's robustness to transmission loss is briefly discussed in section 5. Results from the simulated system are given in section 6.

2. SUB-BAND ANALYSIS AND SYNTHESIS

The technique of sub-band coding can be briefly explained as follows: The input signal is passed through a bank of band pass filters, the analysis filters. Due to the reduced bandwidth, each resulting component may be sub-sampled to its new Nyquist frequency, thus yielding the sub-band signals. Following that, each sub-band would be encoded, transmitted and, at the destination, decoded. To finally reconstruct the signal, each sub-band is up-sampled to the sampling rate of the input. All up-sampled components are passed through the synthesis filter bank, where they are interpolated and subsequently added to form the reconstructed signal. This is depicted in fig. 1 for the case of two bands. With separable filters, multi-dimensional analysis and synthesis can be carried out in stages of uni-directional filters. For example, an image might first be analyzed in its vertical direction followed by horizontal analysis.

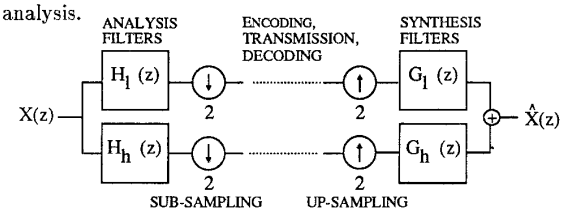


Figure 1. A two band sub-band coding system.

It is desirable to chose the filters $H_l(z)$, $H_h(z)$, $G_l(z)$ and $G_h(z)$ so that $\hat{X}(z)$ is a perfect, but possibly delayed, replica of the input signal $X(z)$, in the absence of coding and transmission loss. We refer to this as perfect reconstruction, which can be written as

$$\hat{X}(z) = z^{-k}X(z), \qquad k \in \mathcal{N}. \qquad (1)$$

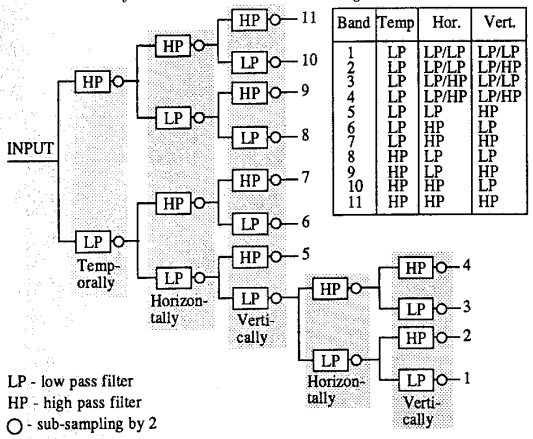
The implementation of a three dimensional sub-band analysis system is shown in fig. 2. The system consists of temporal, horizontal and vertical filtering, with further spa-

CH2561-9/88/0000-1100 $1.00 © 1988 IEEE

tial analysis of the sub-band that has been obtained through low pass filtering in all of the three first stages, which yields bands 1-4. This repeated sub-band division is necessary in order to achieve high compression. Due to the sub-sampling the bit rate in each of the output branches from an analysis stage is half of its input rate. Consequently, the first stage with temporal filters has the highest computational burden since there is no parallelism, while each of the spatial filters operate in parallel at a lower rate. Hence, we chose the temporal filters to be the shortest possible, which also minimizes the number of frames needed to be stored as well as the delay. In the z-transform domain the filters are

$$H_l(z) = (1 + z^{-1})/2, \qquad H_h(z) = (1 - z^{-1})/2. \quad (2)$$

Together with synthesis filters given by $G_l(z) = -2 \cdot H_h(-z)$ and $G_h(z) = -2 \cdot H_l(-z)$, these filters yield perfect reconstruction as in (1), with the delay $k = 1$. Temporal filtering allows us to exploit the vast amount of redundancy in the time dimension without resorting to interframe prediction, which is usually not as robust to transmission error. Note that the decomposition over time yields one component with essentially constant bit rate when encoded, while the other exhibits a bursty behavior after encoding.



Band	Temp	Hor.	Vert.
1	LP	LP/LP	LP/LP
2	LP	LP/LP	LP/HP
3	LP	LP/HP	LP/LP
4	LP	LP/HP	LP/HP
5	LP	LP	HP
6	LP	HP	LP
7	LP	HP	HP
8	HP	LP	LP
9	HP	LP	HP
10	HP	HP	LP
11	HP	HP	HP

Figure 2. Three dimensional sub-band analysis.

The spatial filters operate in parallel at a lower rate, and the filter lengths do not affect the storage requirements. Consequently, we can allow longer filters. LeGall has derived pairs of perfect reconstruction filters which are well suited for image processing [4]. From among them we have chosen the following filters for the spatial sub-band analysis:

$$H_l(z) = (-1 + 2z^{-1} + 6z^{-2} + 2z^{-3} - z^{-4})/4, \quad (3)$$
$$H_h(z) = (1 - 2z^{-1} + z^{-2})/4. \quad (4)$$

When the synthesis filters are given by $G_l(z) = -H_h(-z)$ and $G_h(z) = -H_l(-z)$, the input/output relationship can be shown to be given by (1), with the delay $k = 3$. These filters have linear phase, relatively good characteristics for frequency selection and interpolation, and low computational complexity. The latter aspect is studied and compared to other methods in section 4.

For the purpose of parallel implementation, we will treat the sub-bands as being independent. This is an approximation because the filters have overlapping frequency regions, and the signal statistics in general yield dependent sub-bands. Note that seven of the bands each have only 1/8 of the input data rate, and the remaining four 1/32 of the input rate. Thus, the encoding can be performed at a lower rate.

3. ENCODING OF THE SUB-BANDS

The sub-band analysis has not resulted in any compression: the sum of data points in the sub-bands equals that of the input. However, it has yielded an attractive separation of the data: band 1 has an intensity distribution similar to that of the input, while bands 2 - 11 have distributions highly concentrated around zero with highly reduced variance compared to that of the input. These ten bands are therefore PCM encoded since little can be gained by reducing the already low amounts of correlation through predictive coding. Band 1 is still highly correlated in all dimensions, which is partially removed by one dimensional DPCM encoding. Quantization of pixel values and prediction error respectively, results in large connected areas of zero-valued data points in all bands. This is exploited by run length coding the locations of the non-zero values in each quantized band.

The quantizers for all 11 bands were designed similarly: a wide zero-level to eliminate low intensity picture noise [2], symmetric quantization with uniform steps, and virtually no upper limit on the quantizer. Thereby only two parameters have to be adjusted to fit the quantizer to the sub-band: the width of the zero-level, and the step-size. The output levels are taken as the mid-point between two quantization levels, and they are represented by a set of variable length code words fitted to an exponential distribution. The quantizers outer limit is decided by the point where the variable length code words become unpractically long, e.g. 10 bits. It is beneficial to allow the "outliers" of the distributions to be transmitted, since they yield negligible increase in bit rate, but alleviate some highly visible distortion, like instances of slope overload in the DPCM case.

One advantage with sub-band coding is that even though the quantizers of the sub-bands are fixed, the sub-band analysis has reduced the variance and has divided the signal into bands of varying importance. Since each quantizer is adjusted to the particular sub-band upon which it operates, an adaptivity to the data has thus been achieved. However, the sub-band analysis also offers the framework of doing highly adaptive quantization. For example, band 8 is well suited for estimating motion in the scene. This information can be used to reduce the spatial resolution of the temporal high frequency bands during substantial motion. Analogously, the temporal resolution of low temporal frequency bands (1-7) can be lowered during low motion, while refining the spatial quantization. Also, the quantization of the bands 2-11 may be adapted according to the intensity level of the low pass filtered component, band 1. In accordance with Weber's law, this enables the quantization to be coarser in high intensity areas than in low intensity regions, without causing graininess in the reconstructed image. The linkage introduced between the bands complicates a parallel implementation, but the amount of data to be communicated is small, making such an implementation still practical.

4. COMPLEXITY AND PARALLELISM

The computational complexity of the above described sub-band coding is low: the temporal filters each requires only one addition and one shift operation per value; the length 3 spatial filters require each 2 additions and 2

shifts per value, and the length 5 filters need 5 additions and 3 shifts each. We compare the scheme, as described in section 2 and 3, with two other three dimensional video coding methods: discrete cosine transform coding with interframe prediction, and intra/interframe DPCM. A comparison to vector quantization has not been included since its complexity cannot be easily compared to that of the above methods. The computations pertain to the analysis, the transform, and the prediction for the respective schemes. Hence, we assume the quantization to be of comparable complexity for all three methods. Although not included, motion compensation and entropy coding can be applied to all the considered schemes. The complexity has been calculated based on the following criteria:

i. Sub-band coding: the implementation shown in fig. 2, with filters given by (2)-(4). A filtering is only performed for the values which are going to be retained in the subsequent sub-sampling. The number of operations include two additions per pixel for the DPCM encoding performed on band 1. The complexity of the sub-band decoder is similar to that of the encoder, since every second sample to be synthesis filtered is zero, whereby the effective filter length is halved.

ii. Discrete cosine transform with interframe prediction: DCT with blocks of size 8×8 and 16×16, where the number of operations is taken from [1]. The transform is followed by interframe prediction, which, per transform coefficient, requires one subtraction for the prediction and one addition for replenishment of the prediction loop. The operations can be either floating point, or the transform vectors can be scaled to a suitable precision of integer arithmetic.

iii. Intra/interframe DPCM: the complexity is directly proportional to the order of the predictor used. We chose the following predictor: $\hat{e} = x[t,i,j] - (a_1 x[t,i-1,j] + a_2 x[t,i,j-1] + a_3 x[t,i-1,j-1] + a_4 x[t-1,i,j] + a_5 x[t-1,i-1,j] + a_6 x[t-1,i,j-1] + a_7 x[t-1,i-1,j-1])$, where $\hat{e}$ is the prediction error and $x[t,i,j]$ is the pixel in frame t, row i, and column j. We consider two cases: the optimal, where all seven coefficients are floating point and possibly different; and a sub-optimal case, where $a_4 = 1$, $a_1 = a_2 = a_5 = a_6 = 1/2$, and $a_3 = a_7 = 1/4$. In addition to the prediction, there is one operation to add the quantized value back to the prediction loop. Depending on the coefficients a_i, the operations can be floating point, or integer arithmetic, for which some multiplications may be carried out as shift operations.

The number of operations per pixel of the input image is tabulated for the three methods in table 1. The sub-band coding has a complexity which compares favorably to the other schemes. In terms of storage requirement, both the sub-band coding and the transform/interframe coding require two frames to be stored. The DPCM scheme may operate with one frame and two scan-lines stored; where the frame is stored in the prediction loop and the two scan-lines are the memory needed for the input frame.

In terms of parallel implementation, the structure of the sub-band coder shown in fig. 2 clearly indicates how the data-flows branch out after each stage of filters, where the rate in each branch has been reduced by a factor 2. The tree structure of fig. 2 may be too complex for a hardware architecture. In that case, the scheme can be made completely parallel by cascading the filters that come in sequence and subsequently sub-sample in all dimensions at the end. This would yield 11 parallel and independent three dimensional filters, for which the computations can be organized in such a way that there is no increase in arithmetic complexity. The DCT based scheme can also be implemented in parallel since each block of pixels is processed independently of all others. In contrast to these two schemes, the DPCM is not well-structured for parallel implementation. The prediction gain decreases with each separation of the data. For example, a separation of interlaced video signals to be processed as two separate fields would perform well during periods of high motion in the captured scene, but the prediction would not be the best possible during low motion. When considering parallel architectures, the computational complexity for each of the parallel branches is of interest. The sub-band analysis tree in fig. 2 is unbalanced, with the longest branch for band 1. The sequential complexity is 2.9 additions and 1.9 shifts per pixel in the input, including the DPCM encoding. Consider the coding of a sequence with, say, 15 frames per second of size 512×480 pixels, as used in section 6. The sub-band coding would require maximally 10.7M additions and 7.0M shifts per second.

METHOD	ADDS	SHIFTS	MULTS
i	8.9	6.6	0.0
ii 8x8	8.5	0.0	4.0
ii 16x16	11.2	0.0	5.5
iii Opt.	8.0	0.0	7.0
iii Sub-opt.	8.0	3.0	1.0

Table 1. Operations per pixel for coding methods i-iii.

5. ROBUSTNESS

Recall that the application of the coding scheme is to transmit video over packet-switched networks. When the transmission over such a network is lossy, packets of data are lost, each typically of size 1024 bits. The robustness pertains to the perceived quality of the reconstructed video signal when such loss occurs. The most important issue in alleviating the effects of lost packets is to restrict the propagation of the error beyond the lost values by avoiding dependencies between data in different packets. If the lost data can be treated as erasures, an appropriate substitution can be made. For bands 2-11, the erasures are replaced by zeroes. This makes the loss of a packet appear as lost detail in the reconstructed image. However, we have found that the effect is too subtle to be noticed when the reconstructed sequence is viewed at video rate, even when 3% of all packets in bands 2-11 are lost. For the DPCM encoded band, temporal or spatial interpolation can be used to create replacement for the erased data. In a first attempt the replacement was taken from the corresponding area in the previous frame, which did not perform satisfactory in regions with motion. Consequently, we have initiated a study to use motion compensation to find the appropriate area of the previous frame to be used as replacement. Note that error correcting codes cannot be used alone: the network can usually not guarantee a bound on the packet loss, thus, the loss may exceed what is correctable by the code.

6. RESULTS

We have compressed a monochrome video sequence consisting of 50 frames of size 512 × 480 pixels, with a video rate of 15 frames/second. Thus, the corresponding input rate is 29.5 Mbit/s, which was compressed to an average rate of 1.6 Mbit/s, with standard deviation 416 kbit/s. The bit rate of the encoded signal is plotted in fig. 3. Note that due to the temporal sub-sampling the frame rate is halved. Plotted are also the total rates of the bands with low and high temporal frequencies respectively. Clearly, the temporal filtering gives one component with nearly constant rate, while the other is highly varying. The associated statistics for bands 1-7 are: the mean rate 835 kbit/s, with standard deviation 63 kbit/s; and for bands 8-11, the mean rate is 761 kbit/s, with standard deviation 375 kbit/s. The minimum SNR for any single frame is 35.9 dB, calculated as $20 \cdot log_{10}(256/\sigma_{diff})$, where 256 is the intensity range of the input and σ^2_{diff} is the variance of the difference between input and output frames. The mean SNR is 36.9 dB, and the standard deviation is 0.64 dB. In fig. 4 a-d, one frame from the video sequence is shown for the cases: (a) the original, (b) coded with lossless transmission, (c) coded with 5 lost packets, of size 1024 bits, randomly distributed among bands 2-11; and (d) the previous case with the addition of one lost packet in band 1, which has been replaced by the corresponding area in the previous video frame.

7. CONCLUSIONS

We have described a video coding scheme based on three-dimensional sub-band coding, to be used in a packet-switched environment. In our study the scheme has revealed appealing properties, such as high compression with good perceptual quality, low computational complexity, architectural simplicity for parallel implementation, and relatively good robustness to packet loss.

ACKNOWLEDGEMENTS

The authors wish to thank Bell Communications Research for providing the image sequence, Dr. Didier LeGall and Dr. Ali Tabatabai for helpful discussions. This work was supported by the National Science Foundation under grant CDR-84-21402.

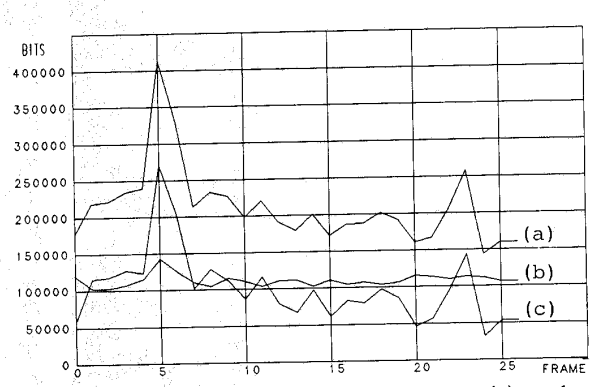


Figure 3. Bit rates of (a) the total output, (b) bands 1-7, and (c) bands 8-11.

REFERENCES

[1] W-H Chen et al., "A Fast Computational Algorithm for the Discrete Cosine Transform," IEEE Trans. on Comm., Vol. COM-25, No. 9, Sept. 1977, pp. 1004-1009.

[2] H. Gharavi and A. Tabatabai, "Application of quadrature mirror filtering to the coding of monochrome and color images, " *Proc. of ICASSP-87*, Dallas, April 1987, pp. 2384-2387.

[3] G. Karlsson and M. Vetterli, "Sub-band Coding of Video Signal for Packet-Switched Networks," *Proc. of SPIE Conf. on Visual Communications and Image Processing II*, Vol. 845, Cambridge, MA, Oct. 1987, pp. 446-456.

[4] D. J. LeGall, "Sub-band Coding of Images with Low Computational Complexity," *Picture Coding Symposium*, Stockholm, Sweden, June 1987.

[5] M. Vetterli, "Multi-Dimensional Sub-Band Coding: Some Theory and Algorithms," Signal Processing, Vol. 6, No. 2, Feb. 1984, pp. 97-112.

[6] J. W. Woods, and S. D. O'Neil, "Sub-Band Coding of Images," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-34, No. 5, Oct. 1986, pp. 1278-1288.

Figure 4. (a) The input, (b) coded, (c) coded, with lossy transmission in bands 2-11, and (d) coded, with transmission loss in all bands.